UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

Evaluation of Image Registration Performance and Study of Classification Algorithm on Histopathological Images

Jessica Monteiro Luzio

Mestrado Integrado em Engenharia Biomédica e Biofísica
Perfil em Sinais e Imagens Médicas

Versão Definitiva - Pública

Dissertação orientada por:
Professor Doutor Nuno Miguel de Pinto Lobo e Matela
Doutor Oliver Grimm

2018

“It always seems impossible until it’s done.”

- Nelson Mandela

Acknowledgements

First of all I want to express my deepest gratitude to Dr. Fabien Gaire, head of Pathology & Tissue Analytics, who provided me the opportunity to carry out this internship under the best conditions. Thank you for this exceptional experience.

To my supervisor Dr. Oliver Grimm and to Dr. Claudia Ferreira, I want to say that I could not have been luckier. Special thanks for your continuous guidance and encouragement. For the great discussions that made me continue and always try to do better. My gratitude also to my internal supervisor, Dr. Nuno Matela, for always being very helpful.

I am grateful to the entire pRED team for providing the best work environment. Thank you for your support, for all the good times and delicious cakes. It was a pleasure to work with such a great team. Special thanks to Eldad Klaiman and Amal Lahiani. Eldad, for you I have to express my immense gratitude and my special appreciation. Thanks for sharing your knowledge and for all the good times during these 9 months. And Amal, thanks for your encouragement and support everyday. Thanks for giving valuable inputs for my work. Thanks for your friendship and for all the joyful times that made work easier. Natalie Zwing, Marc Zaetschk, Konstanty Korski and Hadassah Sade, thank you for your support.

Camille, it was a pleasure meeting you! Thank you for your support during the 6 months together. Thank you for always being there and for showing me that it is not that bad to be far from home.

I also want to thank all the Portuguese and Portuguese speaking people that I met at Roche. It was a pleasure to meet you! Despite being abroad you made me feel home.

Thanks to all my friends and colleagues. You know that your support and encouragement was worth more than I can express on paper.

My acknowledgement will never be complete without the special mention of the one who has been not only the best colleague but also the best friend since day one. Rui, thank you for being a great companion, for all the support, encouragement and friendship. We have seen and helped each other through both good and bad times. The conclusion of these 5 years would never have been possible without you.

And finally, last but by no means least, to my family. I deeply thank my parents, sister and grandparents who have provided me through moral and emotional support in my life. Thank you for your unconditional trust, encouragement, and endless patience. Thanks for being my main support in all moments, specially in the most challenging ones. Words will never be enough to thank you for everything you have done for me.

Contents

List of Figures

List of Tables

Abbreviations

αSMA	α-**S**mooth **M**uscle **A**ctin
ATD	**A**utomatic **T**issue **D**etection
CCC	**C**oncordance **C**orrelation **C**oefficient
CD	**C**çuster of **D**ifferentiation
CRC	**C**olo**r**ectal **C**ancer
DAB	**Di**a**m**ino**b**enzidine
DTA	**D**iscovery and **T**ranslational **A**rea
FFPET	**F**ormalin **F**ixed **P**araffin **E**mbedded **T**issue
FOV	**F**ield **O**f **V**iew
GT	**G**round **T**ruth
H&E	**H**aematoxylin and **E**osin
HTX	**H**aema**t**o**x**ylin
ICA	**I**ndependent **C**omponent **A**nalysis
IHC	**I**mmuno**h**isto**c**hemistry
kNN	**k** **N**earest **N**eighbor
LR	**L**ogistic **R**egression
PCA	**P**rincipal **C**omponent **A**nalysis
PTA	**P**athology and **T**issue **A**nalytics
pRED	**p**harma **R**esearch and **E**arly **D**evelopment
RF	**R**andom **F**orest
SBS	**S**equential **B**ackward **S**election
SFBS	**S**equential **F**loating **B**ackward **S**election
SFFS	**S**equential **F**loating **F**orward **S**election
SFS	**S**equential **F**orward **S**election
SMC	**S**mooth **M**uscle **C**ells
SMOTE	**S**ynthetic **M**inority **O**versampling **T**echnique
SVM	**S**upport **V**ector **M**achine
WSI	**W**hole **S**lide **I**mage

To my sister and parents. Thank you.

Abstract

The project described in this thesis was developed in the Roche Pharma Research and Early Development (pRED) department in Penzberg, which is focused on development of personalized healthcare strategies. The primary goal of the study is to explore computational tools used in Digital Pathology, more specifically on image registration and classification.

Digital Pathology refers to an image-based environment that enables the conversion of glass slides into digital images, and also the management and interpretation of it, involving an enormous amount of data to be handled. The image analysis tasks are fundamental since they allow the identification and classification of different types of structures on tissue samples. Staining techniques as Hematoxylin and Eosin (H&E) and Immunohistochemistr (IHC) techniques are used, allowing the extraction of relevant information for the study of pathological processes. Image registration tools to transfer the annotations from one slide to other are important and routinely used in image analysis processes. Image classification is also a key task to identify and distinguish different type of structures in the tissue.

The first part of this project investigates the feasibility of an automated method for image registration assessment whereas the second part studies the influence of classifiers, features and training set changes in the classification performance. With this we aim to gather information for further improvements on the process of image analysis.

After applied the registration algorithm, a distance based method was developed to evaluate its performance. Distances between each point on target and source annotations were calculated and summed up in the median value. The results of automatic annotations were compared with the manual, in order to ascertain if they behave the same way and so, if the automatic approach can be used for performance evaluation. The assumption that the closest points between two annotations are the correspondent tissue points, makes the results to not give valid information when the registration is not good. As the comparison is done between tissue outline, even when the two annotations are misaligned, there are still close points between the two, even though they do not correspond to the same tissue parts. This approach is then not reliable to use for performance assessment.

Regarding image classification, an algorithm for vessels detection is used and three different settings are tested in order to identify the influence on classification performance: classifiers, features and training set. For the classifiers no significant differences were identified. It was also verified that increasing the number of features does not necessarily increase the classification performance. With regard to the

training set, for one of the ground truth provided, a good improvement was verified when balancing the data set. However, compared with information provided by a different pathologist, the results were not improved with the balancing. This can also be due to the undersampling done for balancing, which might have discarded useful samples.

Although this study may not provide sufficient information to draw general conclusions, it is a starting point for future studies that may improve what was done so far.

Keywords: Digital Pathology; Histopathology; Image Resgistration Performance; Image Classification; Image Analysis

Resumo

O projeto descrito nesta tese foi desenvolvido no departamento dedicado ao desenvolvimento de estratégias de cuidados de saúde personalizados da Roche, em Penzberg, Alemanha. Desenvolvendo produtos nos varios ramos da medicina, esta empresa tem a oncologia como uma das suas areas de maior destaque.

O diagnóstico médico depende muito da avaliação histopatológica do tecido do paciente. Com base na exploração microscópica de amostras de tecido de pacientes, os patologistas confirmam a presença da doença, bem como o seu grau, e orientam as decisões clnicas. No entanto, este método convencional não só consome tempo como também introduz subjetividade e variabilidade no diagnóstico, intrínseco à natureza humana.

Além disso, a disponibilidade de grandes quantidades de dados de pacientes e a necessidade por parte dos laboratórios de histopatologia no processamento de amostras em períodos recordes de tempo, contribuíram para o desenvolvimento de soluções digitais que visam melhorar a reprodutibilidade e a velocidade de analise e complementar as decisões dos patologistas.

A Patologia Digital refere-se a um conjunto de processos que permite a conversão de lâminas de vidro com amostras de tecido em imagens digitais, bem como a gestão e interpretação das mesmas, envolvendo assim uma enorme quantidade de dados a serem manipulados. As tarefas de análise de imagem são fundamentais pois permitem a identificação e classificação de diferentes tipos de estruturas em amostras de tecidos. Para isso, são utilizadas técnicas de coloração como Hematoxilina e Eosina (H&E) e Imunohistoquímica (IHC), que permitem a extração de informações relevantes para o estudo de processos patológicos.

Para começar a desenvolver um algoritmo de analise de imagem, todos as imagens do estudo devem ser anotadas. Este processo consiste em delimitar as diferentes areas do tecido e é feito por um patologista para todas as secções do bloco de tecido em estudo. Esta é uma tarefa naturalmente demorado e, por isso, métodos para a tornarem mais expedita são utilizados.

O registo de imagem é uma ferramenta utilizada para transferir as anotações de uma imagem anotada para a imagem consecutiva. É então essencial para uma boa analise, que o registo tenha uma boa performance. Um mau desempenho nesta tarefa pode levar a uma analise errada do tecido, uma vez que a distribuição espacial dos resultados esta comprometida. A avaliação do desempenho do registo de imagem é por si

só uma tarefa desafiadora e manual e, por isso, existe a necessidade de definir métodos quantitativos e automaticos para o efeito.

Outra tarefa fundamental no processo de analise de imagem é a classificação, que permite identificar e distinguir diferentes tipos de estruturas no tecido. Para gerar os modelos de classificação, um conjunto de dados de entrada e respetivas respostas conhecidas são utilizadas para treinar um modelo que gere previsões razoaveis para a resposta a novos dados desconhecidos. Neste contexto, os dados são imagens com estruturas de interesse para classificar. De forma a chegar a um modelo com a performance necessaria, não só o conjunto de dados de treino pode ser ajustado como também os diferentes classificadores e *features* podem ser testados.

Assim sendo, a primeira parte deste projeto investiga a viabilidade de um método automatizado para avaliação de registo de imagens, enquanto a segunda parte estuda a influência de diferentes classificadores, características e dados de treino, no desempenho da classificação. Com isto, pretendemos reunir informações para posteriormente incluir em melhorias do processo de análise de imagens. Relativamente ao registo de imagem, o objetivo é definir métodos quantitativos para avaliar o desempenho, que poderão posteriormente ser utilizados para avaliação rotineira do registo em imagens histopatológicas.

Após aplicado o algoritmo de registo, foi desenvolvido um método baseado em distâncias para avaliar a sua performance. As distâncias entre cada ponto das anotações de destino e de origem foram calculadas e resumidas pelo valor mediano. Os resultados das anotações automáticas foram comparados com os das anotações manuais, a fim de verificar se seguiam o mesmo padrão e, portanto, se a abordagem automática pode ser usada para avaliação de performance. A suposição de que os pontos mais próximos entre duas anotações equivalem aos pontos de tecido correspondentes, faz com que os resultados não forneçam informações válidas quando o registo não é bom. Como a comparação é feita entre o contorno do tecido, mesmo quando as duas anotações estão desalinhadas, ainda existem pontos próximos entre si, mesmo que não correspondam às mesmas partes do tecido. Assim sendo, esta abordagem não é confiavel para uso na avaliação de performance.

Adicionalmente, foi também explorada a influência da dist¢ncia fsica entre secções de tecido e o desempenho do registo. Para isso foi assumido que o valor da mediana das dist¢ncias seria um bom indicador de performance. Os resultados obtidos são satisfatórios e de acordo com o esperado. O desempenho do registo é melhor para as secções mais próximas da secção alvo de registo (isto é, o valor da mediana das dist¢ncias aumenta com o afastamento das secções). Ainda assim o aumento é muito pequeno, o que significa que apesar da performance piorar com o aumento da distância física, este aumento não é significativo.

Em relação à classificação de imagens, o estudo tem como objetivo estudar a influłncia destes diferentes par¢metros no desempenho da classificação. O objetivo é estabelecer algumas regras para orientar e apoiar o desenvolvimento de futuros algoritmos. Um algoritmo para deteção de vasos foi utilizado e três parâmetros diferentes foram testados (classificadores, *features* e conjunto de treino), a fim de identificar a influência no desempenho da classificação. Para os classificadores, não foram identificadas diferenças significativas. Verificou-se também que aumentar o número de *features* não aumenta necessariamente

o desempenho da classificação. Em relação ao conjunto de dados treino, uma melhoria foi verificada ao equilibrar o conjunto de dados. No entanto, o mesmo não de verificou ao comparar com os dados fornecidos por outro patologista. Esta situação pode também dever-se à subamostragem feita para tornar as diferentes classes mais equilibradas, o que pode ter descartado amostras úteis. Embora este estudo possa não fornecer informações suficientes para tirar conclusões gerais, é um ponto de partida para estudos futuros, que podem melhorar os métodos e análises abordados nesta tese.

Existem ainda pontos que podem e devem ser melhorados para que conclusões confiaveis possam ser tomadas em relação ao desempenho da classificação. Primeiro, mais algoritmos de classificação devem ser usados. Usando apenas um classificador para um tipo especfico de classificação não fornece resultados suficientes para a generalização das conclusões. Para além disso, mais testes para cada estudo (classificadores, *features* e dados de treino) devem ser feitos.

Em segundo lugar, uma abordagem diferente pode também pode ser tomada em relação aos dados usados. Para este algoritmo os dados de validação consistiam na contagem de cada dos vasos de cada classe em cada segmento de imagem. No entanto, tendo as posições de cada vaso permitiria uma analise mais profunda, incluindo, por exemplo, uma matriz de confusão (verdadeiros e falsos positivos e negativos). Além disso, para testar com maior precisão a influência dos dados de treino na classificação, métodos de amostragem como o SMOTE proposto por Chawla [17], poderiam ser aplicados, criando amostras artificiais da classe menos representada.

É relevante notar que este tipo de análises, tanto para avaliação de registo como para classificação, são ferramentas importantes para os patologistas que lidam diariamente com imagens histopatológicas. Tendo métodos computacionais confiaveis permite-lhes economizar o tempo na avaliação visual de imagens, permitindo que as suas horas possam ser dedicadas a tarefas mais complexas.

Palavras-Chave: Patologia Digital; Histopatologia; Registo de Imagem; Classificação de Imagem; Análise de Imagem

Preface

Doing now what patients need next: this is the purpose of Roche, one of the worlds great research-based healthcare companies. It was founded in 1986 in Basel, Switzerland, and it is active in more than 150 countries in all continents.

Roche develops products worldwide in many branches of medicine, with oncology being one of the most highlighted. With sites all over the world, in Penzberg, greater Munich, Germany, lies one of the largest biotechnology centres of Europe, where both pharmaceuticals and diagnostics divisions are present, with research, development and production.

The figure below shows an aerial view of Roche site in Penzberg.

FIGURE 1: Aerial view of Roche site in Penzberg.

This project was developed in the Roche Pharma Research and Early Development (pRED) department, which is focused on development of personalized healthcare strategies. The Oncology Discovery and Translational Area (DTA) is one of its subdivisions and which vision is the development of curative medicines for cancer patients. Within this division, in the Pathology and Tissue Analytics (PTA) department, drug tests on tissue by using immunohistochemistry (IHC) are being conducted. The Imaging

Lab, where I was directly working, has as main function the development of algorithms for IHC image analysis, so that conclusions and decisions can be easily made by pathologists. This digital analysis can successfully be applied since high technological systems of digital pathology are available in the department.

Chapter 1

Introduction

Histopathology is a word formed from Greek roots: *histos* meaning tissue, *pathos* meaning disease and *logos* meaning study. It refers to the detailed examination of biological tissues, in order to observe and assess the appearance of diseases, such as cancer.

Medical diagnosis relies heavily on the histopathological evaluation of patient tissue. Based on the microscopic exploration of patients tissue samples, pathologists confirm the presence of disease as well as its grade, and guide clinical decisions. However, this conventional method not only is time consuming but also introduces subjectivity and variability to the diagnosis, intrinsic to human nature [1,2].

Moreover, the availability of high amounts of patient data, and the necessity of histopathology labs on processing received samples in record time periods has contributed to the development of digital solutions that aim to improve reproducibility and speed, and complement the pathologists decisions.

Digital Pathology refers to an image-based environment that enables the conversion of glass slides into digital images, and also the management and interpretation of it. This involves an enormous amount of data to be handled and requires accurate and fast digital methods for proving an useful outcome.

The primary goal of this thesis is to explore computational tools used for histopathological image analysis, more specifically on image registration and classification.

This introduction chapter includes the motivation for this project followed by the thesis outline.

1.1 Motivation and Objectives

From pathologists perspective, the image analysis tasks are fundamental since they allow the identification and classification of different types of structures, such as vessels and different types of cells, thus allowing a deep understanding of the tissue.

The visualization of tissue characteristics and the expression of biomarkers is commonly achieved by staining tissue sections with Hematoxylin and Eosin (H&E) or IHC techniques, allowing the extraction of relevant information for the study of pathological processes.

For many biological processes, it is important to understand the co-localization and interaction of biomarkers. Therefore, chromogenic multiplexing methods for multiple antigen labelling are used. The relative spatial distributions between biomarkers can be assessed via registration of consecutive slides stained for different biomarkers of interest to obtain an in-silico multiplexed slide. In addition, registration also allows the expert-drawn manual annotations, delimiting for example tumor or necrotic areas, to be transferred to neighboring slides thus reducing the expert time dedicated to this task.

Besides robust registration methods, also suitable performance assessment tools should be found, in order to ensure the viability of the process and its use in an automated and high throughput manner. The visual assessment by domain experts is the oldest method of accuracy estimation and it is still used for performance assessment, but there is an increasing need for efficient quantitative methods.

Literature review shows that the quantitative approaches reported so far are based on distance calculations between landmarks or annotations identified in both source and target images by experts, or are intensity based metrics.

The first part of this project investigates the feasibility of an automated method for image registration assessment on histopathological images. The research questions are the following:

- Is it possible to assess registration accuracy without manual annotations?
- How does the tissue cutting scheme impacts the registration accuracy? Is the distance between the registered and the target slide relevant?

These study aims to gather information about registration performance and thus improve the current lab workflow.

The second part of this project is related with the image analysis itself. Regardless of the amount of biomarkers to be assessed, digital tools are routinely developed for the analysis of the slides. For each assay, an algorithm is developed or adapted to perform one of many tasks, which typically are counting, classification and/or segmentation.

Automated classification is a core technique for digital pathology as it helps to standardize results and avoid variability, giving highly valuable information to pathologists, such as cells densities or distances between some specific cells, which can led to important conclusions in diagnostics.

For this purpose, supervised machine learning algorithms are used to teach computers to learn from experience and so developing predictive models based on both input and output data. These models are created according to user-defined criteria. Changing the size of the data set for training and using

different features or classifiers, are settings which can be handled, in order to influence the performance of the classification algorithm.

Study the influence of these settings in the classification performance is the goal of the second part of this thesis. To serve as a guide for this research, the following questions will be followed:

- How does the classifier algorithm influence the classification performance?
- How does the features selected impact the classification performance?
- Is the data set size an important factor?

The obtained conclusions can be used to guide and support the algorithms development.

1.2 Organization of the thesis

To explore the research questions mentioned above, this thesis is organized in 5 chapters.

Chapter 1 is intended to contextualize the reader about the domain problem and expose in more detail the motivations and objectives of the developed project.

In Chapters 2 are described the main theoretical concepts that support the thesis. More specifically, it takes a general approach to Digital Pathology, going through the technical procedures of tissue preparation and image production, and the steps involved in the process of image analysis.

In Chapter 3 the basic concepts about image registration in histopathology are introduced and the state of the art is reviewed. Moreover, the main methods developed to date to address the problem of image registration performance assessment are described. The methodology proposed in this study and the results are enumerated and discussed.

In turn, the concepts and literature review on machine learning and classification in histopathology are described in Chapter 4, as well as the methods, results and discussion.

Finally, Chapter 5 presents the conclusions, as well as some of the limitations and future perspectives.

Chapter 2

From Tissue Preparation to Image Analysis

Pathology is a branch of medicine that is concerned with the diagnosis of disease. The main objective of histopathology image analysis is to identify morphological features from tissue, that could be associated with pathological or normal patterns. This tissue exploration enables pathologists to confirm the absence or presence of a disease as well as grade or measure the evolution of it, which makes it considered as the ”gold standard” for diagnosing many cancers.

The figure below schematically represents the process between tissue collection and image analysis.

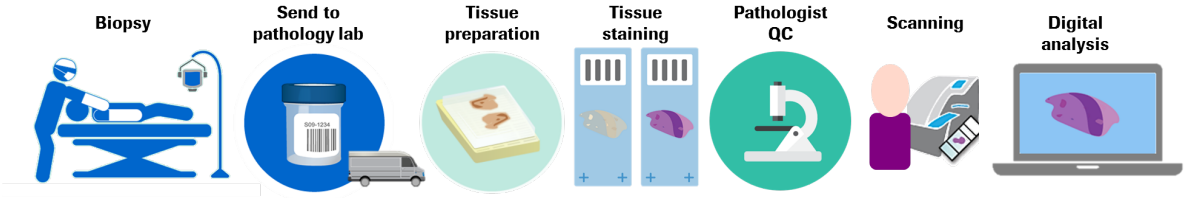


FIGURE 2.1: Schematic representation of the process between tissue collection and image analysis. Adapted from [3].

After receiving the tissue sample, the pathology lab starts with the tissue preparation, which is followed by staining. The stained slides are then reviewed by pathologists to control the quality of the process and, after approval, high-resolution replicas of the glass slides are created using a slide scanner, and uploaded to a database where the images are stored and visualized. These digital representations - known as whole slide images (WSI) - can be computationally manipulated in order to simulate a microscopic examination, and digital image analysis tools can be applied.

In this Chapter, a theoretical overview of the technical procedures involved in the processes from tissue preparation to image analysis will be provided.

2.1 Tissue preparation

Due to the importance of the images in the diagnosis, it is fundamental to ensure high standards during the whole process of tissue preparation to guarantee a reliable analysis. A sequence of technical steps is needed to produce the tissue slices, including fixation, dehydration, clearing, infiltration, embedding and sectioning [4].

After the biopsy, as the tissue is deprived of blood supply, degenerative processes start and substantial changes on the appearance of tissue components arise. Therefore, tissue fixation is the first step, which objective is the preservation of tissue in a "life like state", being one of the most important factors in the production of satisfactory results. On one hand it is important to protect the tissue against the natural degeneration and the effects of the processes to which it will be submitted. On the other hand, tissue must continue reactive to reagents and stains. Heat fixation, perfusion or immersion, are some of the fixation methods used commonly in histopathology. Most fixatives act either by denaturing or precipitating proteins, thereby forming a sponge or meshwork that hold the tissue constituents. Formalin is the most commonly used fixative due to its advantages such as the rapid penetration and the easy availability and good price. By slowly penetrating the tissue, it forms cross links between amino acids and proteins, making them insoluble.

After fixed, due to the importance of cutting thin sections, a sequence of steps needs to be applied to the tissue to provide an adequate support. The first step is the dehydration, using alcohol products to remove the water from the tissue. This alcohols are then cleaned, in the clearing step. After that, the tissue is embedded in a solid and stable medium, such as paraffin wax, which is fluid when hot and solid when cold, providing an external support for sectioning. This forms the so-called Formalin Fixed Paraffin Embedded Tissue (FFPET).

The final step is the sectioning, which refers to the cutting of the embedded tissue into thin sections (e.g. 2.5 μm) which will be placed in glass slides ready for staining.

2.2 Tissue staining

The purpose of the staining process is to reveal the structural details of the tissue specimen, which are mostly colourless. Different dyes are applied to the tissue to highlight different structures and then a cover-slip of glass or plastic is used to protect the specimen.

H&E makes use, as its name suggests, of a combination of two dyes, which are the most popular stains in pathology. Their popularity is due not only to the simplicity in the application but also to the broad range of morphological features they can reveal. With this staining method, structures as nuclei, cytoplasm, muscles and collagen can be differentiated, which give pathologists a general idea of the tissue morphology.

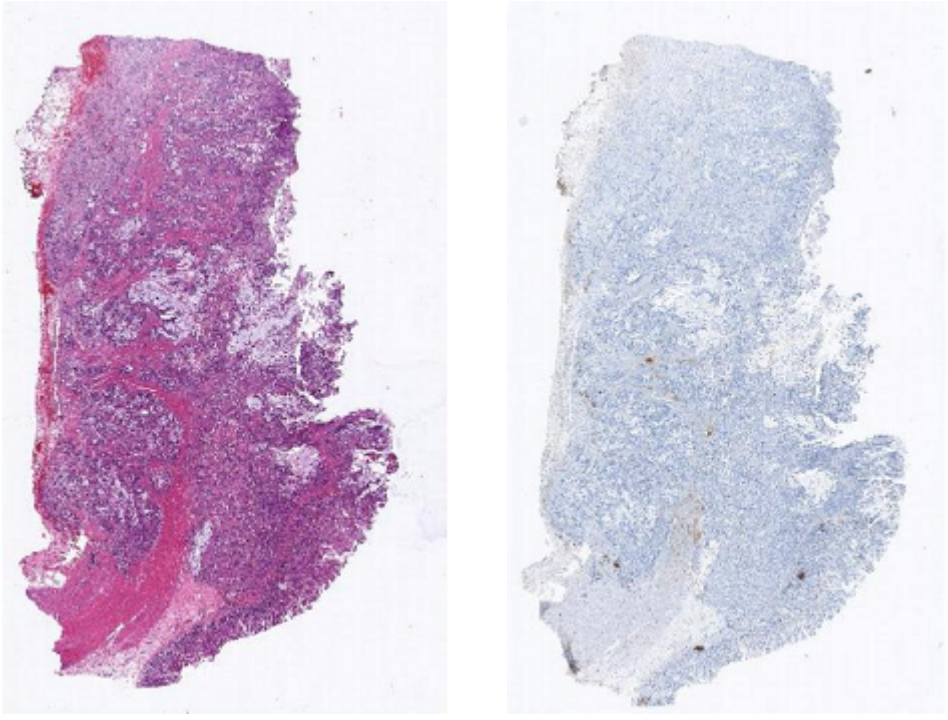
However, for detailed and specific information, more complex staining techniques are needed, such as immunohistochemistry.

IHC refers to a high valuable research method that aims to determine the tissue distribution of an antigen (usually a protein) by using an antibody. The target antigen, also called biomarker, is, in its broadest sense, any biological or physiological entity that aims to assess biological or pathogenic processes, or pharmacological responses to a therapeutic intervention. These biomarkers can be expressed, for instance, by tumour cells.

The process is based on antigen-antibody interactions and in the context of histopathology it remains the gold standard for cancer diagnosis.

The identification and localization of tissue antigens provides valuable information that could not otherwise be obtained through routine H&E stained sections. Even so, this traditional and popular primary staining is still very appreciated by pathologists to study the tissue morphology, which is why it is a common practice to perform H&E staining in some tissue sections, to get a first impression of the tissue organization.

In Figures 2.2(a) and 2.2(b), two consecutive slides are presented with the two staining techniques.



(a) H&E (b) IHC

FIGURE 2.2: Two consecutive slides stained with different staining methods: H&E and IHC. Images retrieved from Roche IRIS platform.

In IHC, the presence of the target antigen can be visualized thanks to sequential process that ends with a chromogenic reaction. First, a solution containing the primary antibody - highly specific for its target - is added to the tissue section. After binding to the antigen, the remaining unbound antibodies are washed away and the second solution is added. This solution contains an enzyme-antibody complex that fixes to the primary antibody and, after enough time, another washing step is followed. Finally, a chromogenic reagent is added and a coloured reaction product is visible by light microscopy. Thus, the target antigen can be localized. This process is for indirect IHC. For direct IHC, instead of using a primary and secondary antibody, just the first is used, which forms the complex with the enzyme (figure 2.3). Normally, after staining the target antigen, a counter-stained is followed, in order to dye the background.

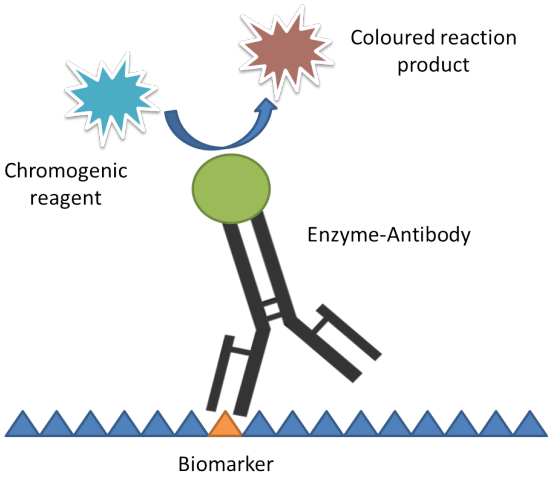


FIGURE 2.3: Schematic representation of the IHC process, using only one primary antibody.

The potentialities of this method led to its growing role in oncology drug research and so new developments have been proposed. In several cases, a single marker is not enough to give answers to the research issues and the localization of different cell types in the tissue specimen is needed. Different markers convey different clinical information and thats why multiplex IHC emerges as a new technique for multiple antigen labelling on the same slide, which means that different biomarkers are labelled with different colours in the visible spectrum. Different chromogenic reagents are commercially available and so, within the range of colours provided, a smart combination should be done in order to distinguish as clear as possible the two antigens. Even so, the coexistence of different labelled biomarkers could be a problem, since the different targeted proteins could be expressed in the same cellular compartment thus leading to a possible colour merging or masking. For visualization, in figure 2.4, three chromogens are represented in the colour space, in a planar representation of 1/8 of a sphere. The better results are achieved with the larger distances between chromogens in the colour space.

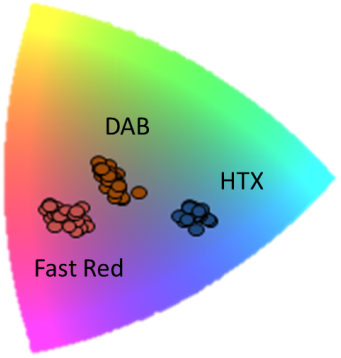


FIGURE 2.4: Planar representation of 1/8 of a sphere showing an example of colour unmixing, using as chromogens Fast Red, 3,3'-Diaminobenzidine (DAB) and Haematoxylin (HTX). Image retrieved from Roche supplementary software.

Stain unmixing algorithms can be applied to overcome this problem and so a WSI is converted into its biomarker-specific images [5]. However, a good colour separation can be challenging, compromising a good comprehensive analysis. To avoid this limitation, a different method can be adopted in order to evaluate each biomarker on a single tissue section. For that, from the same tissue block, consecutive sections are stained with one (or more) different biomarker each. In this situation, complementary image registration procedures are needed in order to group the different results in one single reference slide.

2.3 Digital image analysis

As already mentioned, digital techniques have a powerful impact in histopathological image analysis since a deep quantitative evaluation can be achieved. In the Imaging Lab, the typical workflow that ends up with the analysis of the images is represented in the figure 2.5.

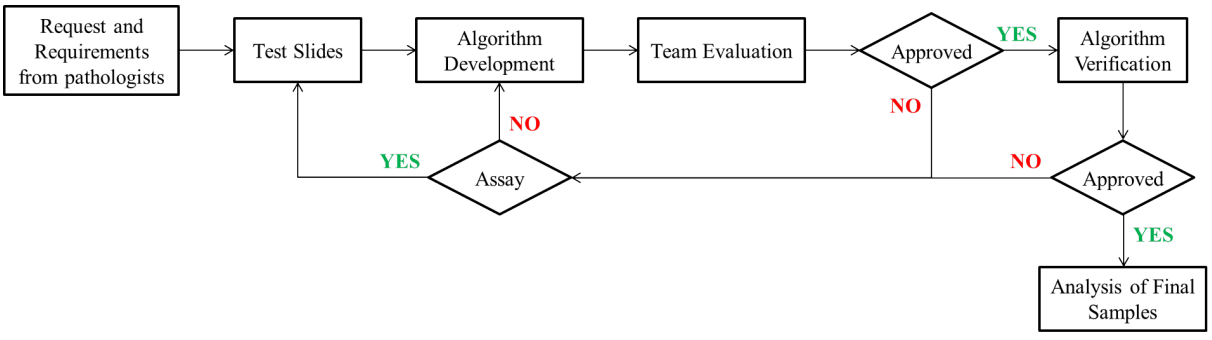


FIGURE 2.5: Imaging Lab workflow for image analysis.

2.3.1 Request and requirements

The process always starts with a request from pathologists. They have specific questions and so they need objective answers. Therefore, to get the intended analysis, several requirements have to be priori set: what they want to detect (e.g. blood vessels stained with DAB chromogen), what they want to be reported (e.g. vessel count, area, intensity of the staining), where they want the detection (e.g. tumour, non-tumour, necrosis) and in which tissue context (e.g. surgical specimens of Ovarian Cancer). Other useful information to construct the algorithm, such as known issues or variability, is also provided. For example, it could be expected to find incomplete staining of some structures or variable staining intensity. Finally, the acceptance criterion is defined. Typically, it is set as acceptable an accordance of 80%-85% between algorithm results and ground truth (GT) provided by pathologists. After collected the needed requirements, the test digitized slides are provided to the imaging team and the development of the algorithm starts.

2.3.2 Algorithm development

The development of an algorithm for image analysis is based on supervised machine learning. Classification techniques are used to create classification models that map input data into categories, to predict discrete responses.

Taking a known set of input data and known responses to the data, a model is trained to generate reasonable predictions for the response to new data. In this context, the data are images with structures of interest to be classified.

To reach this point, the images need a pre-processing that comprises colour deconvolution techniques and object detection. Furthermore, annotations and registration methods have also to be performed during the algorithm development process.

This subsection is organized to follow the general order of these procedures.

2.3.2.1 Annotations

To start developing the algorithm, all the slides of the study have to be annotated. It consists of delimiting the different tissue areas (figure 2.6) and it is done by a pathologist.
The result is a group of labeled pixels corresponding to the outline of the tissue.
The annotations can include excluded areas such as artifacts and tissue losses that will not be analyzed. This process is naturally tedious and time-consuming.

In order to avoid this, registration tools to transfer the annotations from one reference slide to its consecutive are available.

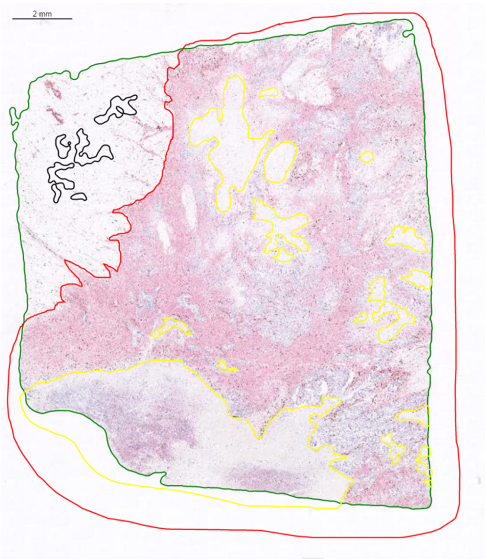


FIGURE 2.6: Tissue annotations drawn by pathologist. Colour legend: tissue (green), tumour (red), necrosis (yellow), exclude (black). Image retrieved from Roche IRIS platform.

2.3.2.2 Registration

This method is not only used to transfer pathologists' annotations between consecutive slides. As mentioned before, they are also needed to group different results in one single reference slide, when each biomarker is used in a different tissue section.

In addition to the visual assessment of populations' distributions, it also allows distance calculations between different cell types, stained in different slides. However, for this purpose the registration results need to be accurate enough. Besides biological factors that cannot be controlled, the accuracy of this registration is strongly linked to the success of the previous registration step, for the annotations.

2.3.2.3 Colour deconvolution

To evaluate the contribution and distribution of each stain, a colour deconvolution step is implemented to separate the different colour channels of the slide.

The method used is based on the one described in [6]. The image is first transformed into the optical density of absorbance space. The Lambert-Beer's law of linear attenuation is used as approximation for this transform (see figure 2.7).

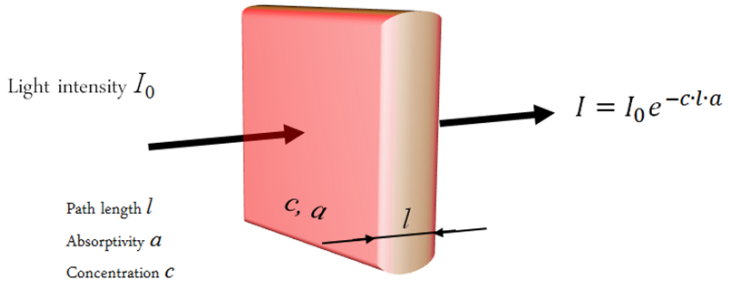
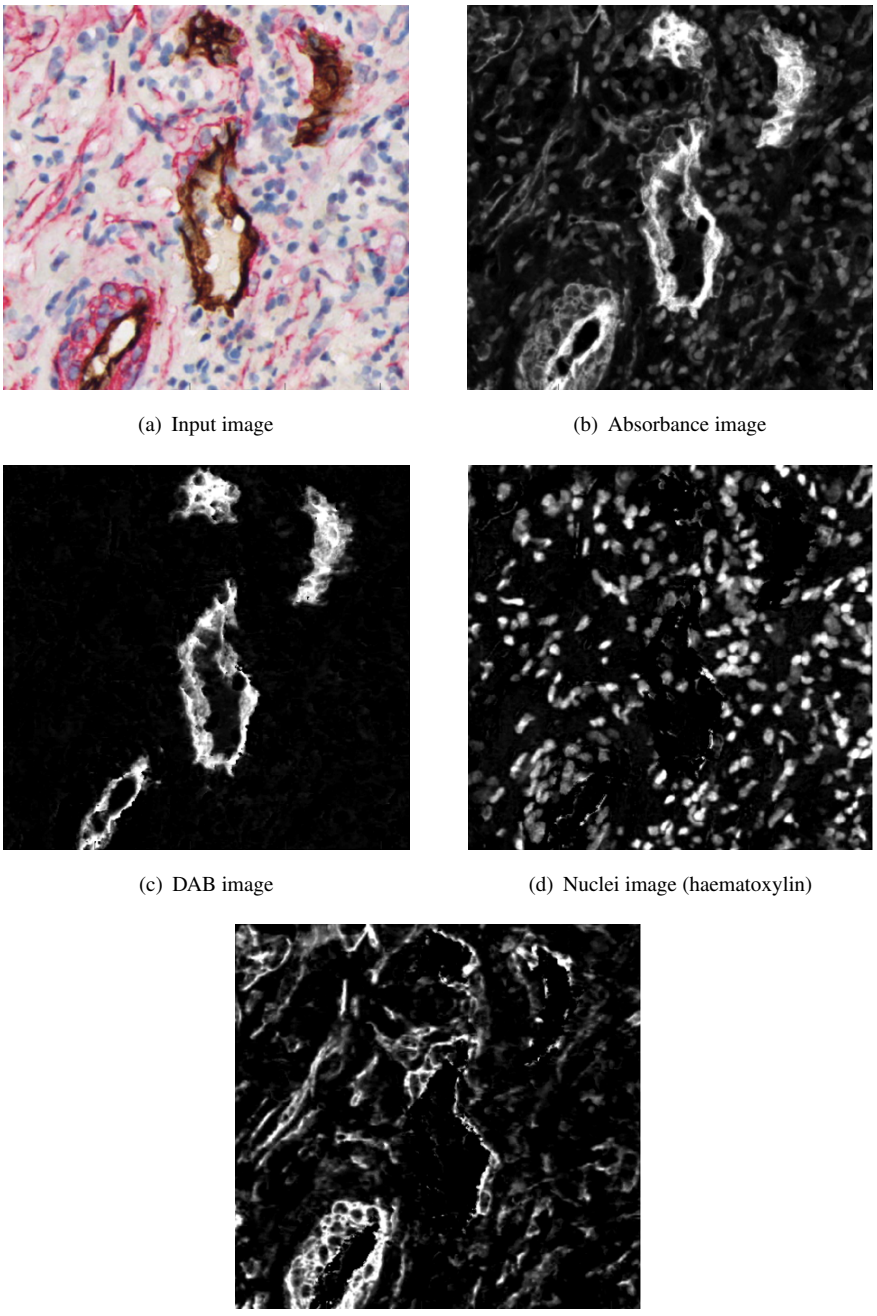


FIGURE 2.7: Illustration of the Lambert-Beers law. Image retrieved from Roche supplementary software.

In this model, light with an incoming intensity I_0 is attenuated by the transition through stained tissue. The linear absorption coefficient is the product of concentration c of the stain in the tissue, path length l through the tissue and absorptivity a of a stain. The exponential absorbance is computed individually for each color channel in an RGB brightfield image.

Therefore, the concentration of each stain can be found and biomarker-specific images are obtained, thus allowing each stain to be analysed individually. In figure 2.8 an example is shown. The images show a field of view (FOV) of a colorectal cancer (CRC) sample. The input image (figure 2.8(a)) was retrieved from Roche IRIS platform. It was used a double staining with DAB and Fast Red chromogens, and haematoxylin as counterstaining. The goal of the brown chromogen is to stain vessels (which express the marker Cluster of Differentiation 34 (CD34)) while the red is for α-smooth muscle actin (αSMA) which can be found in the smooth muscle cells (SMC) of blood vessels. However, Fast Red can also stain other muscle cells in the tissue. Vessels with a wall of αSMA around belong to a class and vessels without belong to other. The remaining staining is of no interest for this classification.



(a) Input image

(b) Absorbance image

(c) DAB image

(d) Nuclei image (haematoxylin)

(e) Fast Red image

FIGURE 2.8: Example of images obtained after colour deconvolution.

2.3.2.4 Object detection

As opposed to the pixel-based representation of computer vision, the humans' concept of the world is based on objects. Thus, it is fundamental to develop computer vision methods to allow an object-level image analysis. Groups of pixels satisfying some similarity criteria define the objects of interest, which

can be vessels, cells, nuclei, etc. To detect and segment these structures both colour and shape characteristics are taken into account. Using essentially threshold, size and shape filters, different structures are segmented based on the corresponding biomarker-specific images.

In figure 2.9 an example is shown, where vessels stained with DAB were segmented. The relation between the image c) in the figure 2.8 and the result obtained is clear.

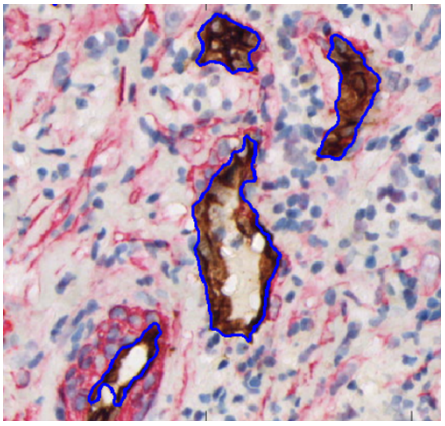


FIGURE 2.9: Result image after vessel detection (blue).

2.3.2.5 User training

The user training is done by pathologists and it is essential to the construction of the model. Pathologists are required to classify the segmented objects to the proper classes thus creating a training data set with the GT. In figure 2.10 it is shown the GT of the FOV represented above, after user training.

This process is performed with FOVs extracted randomly from different WSI. Ideally the training should encompass objects with the greatest diversity of characteristics inside each class. As the name indicates, this set of images is used to train the algorithm.

One of the problems frequently present in many real-world applications is the existence of an imbalanced training set, which means that the classes are not equally represented. Hence, with a poor training of the minority class the accuracy of the classification for this class is usually low.

To avoid that, the training has do be performed in a balance data set and, when this is not possible, solutions have to be found, whether at the data level or the algorithmic level (as discussed in Chapter 4).

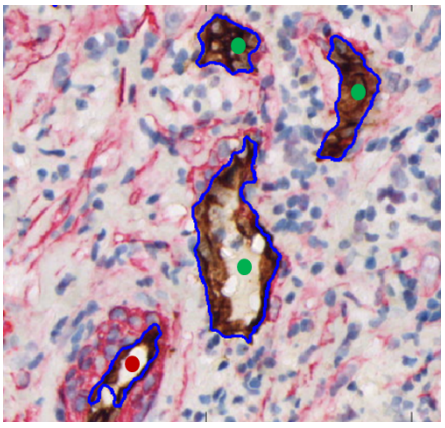


FIGURE 2.10: Result image after user training. Vessels with a green circle belongs to one class and the vessel with red circle belongs to other.

2.3.2.6 Features collection

The goal of the supervised learning is to create a model that predicts the class of each input object according to its attributes. These attributes are called features and are properties intrinsic to the objects themselves and which can be expressed by numbers. A wide range of features can be collected, including gradient, texture, intensity and geometric features, and they can be collected in different relative positions of the object (i.e. inside, outside or in the border). These characteristics are hence used as inputs to separate the different objects in different classes. So, after the user training, it follows the collection of these properties, that should provide informative and non-redundant information for each class.

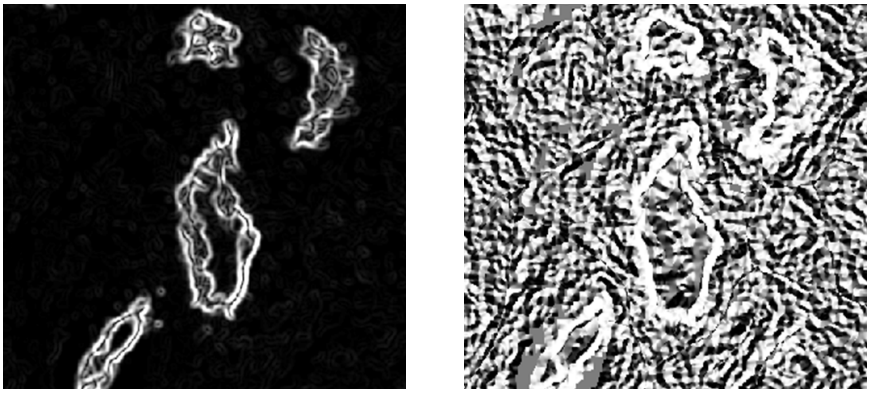
However, extracting features that can describe the different structures and differentiate them perfectly is not possible since there is always variability inside the classes and so the separation is not trivial.

Nevertheless, the most accurate approach should be found and a reasonable number of features should be selected. Although it seems intuitive that the more the better, it is not so trivial. Too many features could led to a too tightly fitting to the training data and also means a much longer process that do not add any value. So, the number of features should be no more than the strictly necessary [7].

To proceed to the collection of these measures, several parameters have to be defined. First, it is necessary to identify what kind of features will be collected (e.g. size, shape); secondly, in which images we want the collection (can be single- or multichannel images); thirdly, in which regions of the objects (e.g. inside, border, from the center to two pixels outside); and finally which descriptive statistics we want to be reported (e.g. mean, standard deviation).

Let's consider the previous figures as an example. As mentioned before, what differentiates the two classes of vessels is the presence or absence of a SMC wall around. Furthermore, it is also necessary to differentiate these vessels (with and without SMC) from the rest of the structures that could eventually be detected in the step of object detection. Thus, the interesting features are mainly related with colour, size and shape, in the border and outside the objects. Since the SMCs are stained with Fast Red and the

vessels with DAB, these biomarker-specific images (2.8(e) and 2.8(c)) are of great interest. In addition, new images (e.g. color transformation, gradient) can be generated in order to get features from them. For example, in figure 2.11 are shown gradient magnitude and direction images generated from the DAB image (2.8(c)). From these images, distances can be measured (shape features) as well as intensities in different relative regions of the structures.



(a) Gradient magnitude (b) Gradient direction

FIGURE 2.11: Example of images obtained after colour deconvolution.

2.3.2.7 Classifiers

To end up with the construction of the algorithm, a classifier has to be chosen. Based on the collected features, a model is created to separate as well as possible the different classes, and so to map the input objects to their categories.

In Chapter 4, a literature review of these classifiers is presented. Their specific characteristics should be analysed in order to do the most accurate choice. The best practice is always to try a bunch of different classifiers and select the best one by cross-validation.

2.3.3 Team Evaluation and Algorithm Verification

2.3.3.1 Team Evaluation

During the development of the algorithm, several test phases are performed to guarantee that the best result is achieved.

Once the algorithm is in the final stage, a team evaluation is followed. Both pathologists and scientists analyse the performance of the algorithm in a qualitative way, in order to decide if the algorithm is already accurate enough to make it worth proceeding to the final and quantitative verification. If the

algorithm does not appear to be accurate enough in this visual assessment, the staining protocol and/or the algorithm have to be reviewed.

This step is performed with a test data set, that is composed of images with the highest variability possible, that are not present in the training set. In this way it is possible to evaluate the performance of the algorithm in images that were never exposed to it.

2.3.3.2 Algorithm Verification

This is the final validation phase in which the algorithm is applied to the validation data set composed by several images that were not used for training. This quantitative analysis aims to compare the results of the algorithm with the GT provided by pathologists and, if the acceptance criterion is achieved, approve the algorithm. Otherwise, the staining protocol and/or the algorithm have to be reviewed and potential changes have to be done. Typically, when changes on the staining protocol (and so on test slides) are needed, they are detected in an early stage and performed before this step. So, in this stage, the non approval of the algorithm typically led to a revision of the algorithm instead of the test slides.

This validation is a major step of the model development. It not only allows to have confidence in the algorithm results but also to eventually record the error margin. This step cannot be disregarded, and its importance is specially high when dealing with patient data in the context of clinical trials.

After analyzed all the results, and after the approval, the algorithm can be applied to the samples of interest, and the questions raised by pathologists at the beginning of the process can be answered.

Chapter 5

Discussion and Conclusion

5.1 Assessment of Image Registration Performance

After presented the methods and results of image registration performance assessment on Chapter 3, we discuss and compare them now with what was expected and already studied by the literature.

The most common methods to evaluate registration are, as already mentioned, based on displacement between landmarks identified by experts (pathologists in this case). As this requires more time invested, the idea of this study was to ascertain if there is a possibility to use a more automated method.

We used then 30 pairs of slides with structures annotated by pathologists and, similar to what is made on the literature, we calculated the distances between the points of each structure. Since we used structures instead of single points, we summed up the distances in one single value (median) for evaluation, which was 5.1 micrometers.

For the same 30 pairs of slides, we run an algorithm of tissue outline detection. This tissue annotations were then used with the same purpose of the manual annotations. Minimum distances were calculated for each point and summed up in the median value. For the 30 pairs of slides the average median obtained was 1.5 micrometers.

Looking beyond the numbers, the boxplots give us a more general understanding of the results obtained. For the manual annotations, several cases have a big range of distances (elongated boxes). For the automatic annotations, the range of distances is smaller and in lower values, however there are more outliers than in the manual cases (observation points that are distant from all the other observations).

These lower values of automatic annotations are easily explained and indicate the reason of not viability of this method. Looking for example for slide pair 10 (represented in Figure 3.6), we verify that the range of distances is high on the manual annotations which indicates that the registration is not good for this case, which leads to a big shift between the source and target annotations.

Even though the bad results are clearly affecting the median on the manual annotations, for the automatic annotated slides the value does not reflect the bad performance. With the assumption that the closest points between two annotations are the correspondent tissue points, the results do not give valid information when the registration is not good. As the comparison is done between tissue outline, even when the two annotations are misaligned, there are still close points between the two, even though they do not correspond to the same tissue parts.

Looking for the median or any other statistical measure will not be enough to provide valid information about registration performance using automatic annotations. A visual assessment and a the use of additional assessment methods (such as boxplots) could be, however, a good combination to have a global overview of performance, using this automatic annotations. The big limitation needs to be always considered, which is the possibility of not be comparing the same tissue points.

Regarding the relation between registration performance and physical distance between source and target slide, the same limitation needs to be considered. For this purpose were used automatic annotations and considered the median values. Assuming that these values reflect the actual registration performance, the results obtained are satisfactory and according to what was expected. The registration performance is better for slides closest to the target slide (median distance increase with increasing slide distance), however the increase is very little, which means that it worsens with increasing physical distance, but not significantly.

5.1.1 Future Work

For this study, different approach could have been followed and should be done for a future and deeper analysis.
Instead of using tissue outline annotations as automatic method, an approach of automatic object detection could have been followed.
With this, the distances could be measured between the automatic structure detection and the manual structure annotation.
Moreover, a more detailed analysis should be performed to understand the outliers results.

5.2 Supervised Machine Learning for Classification

Looking now for the methods and results of Chapter 4, we analyze and discuss what could have been improved in this study.

The algorithm used was already known from the beginning as being a challenge.
Classifying CD34+/αSMA+ vessels is not trivial even for the humans (as suggested in Table 4.2 and Figure 4.1.). Moreover, the small amount of this class comparing with the negative class impacts also the classification result. This is also clear when looking for the overall results of the two classes. Regardless

the combination of features, classifiers or training set, the class CD34+/αSMA- always shows good results, unlike the αSMA+ class.

Regarding the analysis of different classifiers, the differences obtained between the 4 are not relevant differences to be highlighted which confirms what is also in the literature: the best practice is always use simple models that allow a better interpretation. For this study what could be also further tested was the parameters of each classifier which, in this study, were kept as default.

With regard to the feature sets, more conclusions can be taken. Feature set 1 was constructed with 42 features in total. From this set we added then 2 more feature images, which gave the amount of 78 features in the end (Feature set 2). Moreover, as the selection of features pretends to be inspired by visual characteristics defined by pathologists when classifying the structures, for Feature set 3 the range of the outside border was increased, since it is a region which gives important information to pathologists regarding the muscle wall around vessels. For Feature set 4 the idea was to decrease the number of features, ending with a total of 18. Finally, principal component analysis was used in Feature set 5 to perform a linear transformation on the data in order to incorporate 90% of the information in the first few principal components.

For class CD34+/αSMA-, increasing the number of features gives better results, as well as increasing the thickness of the outside border. When looking for feature set 4, which has much less features, the results for pathologist 2 are also better. For pathologist 1, only the percentage of variance explained is 2% less when compared with set 1; all the other measures are better. Comparing the sets 2 (78 features) and 4 (18 features), the second has much less features which gives a reduced complexity to the classification model, without reducing the performance. This could be related with the phenomenon known as the curse of dimensionality. For histopathological image analysis it is fundamental to have a low number of features since images are large and thus the computational complexity would be too high. In this case, 60 features more would only increase computational complexity without any relevant improvement. Using feature set 5, the results are clearly worse.

For class CD34+/αSMA+, the best results are achieved with feature set 4, which is the set with less number of features. Using PCA also improves the results. As mentioned in the last paragraph, this phenomenon is in accordance with what exists in the literature. More features do not necessarily lead to better classification. Contrary to what was expected, increasing the thickness of the outside border did not lead to the expected outcome. For this class we would expect that collecting features in a bigger range of the outside border of the vessel would give a better result. Even so, it gave good improvements in the previous class, probably for the same reasons (identifying that there was not muscle cells around and so the vessel is αSMA-).

For the classification of the vessels in general differences between the 5 different sets tested are not significant. We verify that using PCA the results are worse.

Regarding the training set, for the negative class, and when using as GT the classification of pathologist 1, balancing the training set gave good improvements. For the positive, the best results were obtained with the Training Set 1, although they continue to be far below what is expected of an algorithm for

scientific use. Comparing the algorithm results of total vessels with pathologist 1, what seems to give better results is increasing the training set (Training Set 3). For pathologist 2 the results are better with the first set. Balancing the training set did not improve the results of classification significantly. This could be explained by the method used for balancing: under-sampling. As we removed samples from the most representative classes, useful information might have been discarded.

5.2.1 Future Work

Of course, there is still room for improvements. There are some points that can and should be improved so that reliable conclusions can be taken regarding classification performance.
First, more classification algorithms should be used. Using only one classifier for a specific type of classification does not give enough results for generalization of conclusions. Additional cell classifiers, should be used in the future to increase the amount of results. Also, more tests for each study (classifiers, features and training set) should be done.
Second, a different approach can also be taken regarding the ground truth data. For this algorithm we have for validation the counting of each FOV for each vessels' class. However, having positions of each vessel would allow a deeper analysis including, for example, a confusion matrix (true and false positives and negatives).
In addition, to test more accurately the influence of the training set on classification, methods of over-sampling such as SMOTE proposed by Chawla [17], could be applied, by creating artificial samples from class αSMA+.
Statistical tests should also be applied to get more accurate conclusions.

Finally, it is relevant to note that this type of analysis, for both registration assessment and objects classification, are important tools for pathologists who deal daily with histopathological images. Having a trusted registration assessment methods would save time on image visual assessment, allowing them to dedicate their hours to more complex tasks in which the human being is essential. The same applies for object classification: the conclusions on classification methods lead to the creation of reliable classification models which have direct application on drug tests for example.

Bibliography

[1] Hamilton, P., van Diest, P., Williams, R., & Gallagher, A. (2009). Do we see what we think we see? The complexities of morphological assessment. The Journal Of Pathology, 218(3), 285-291. http://dx.doi.org/10.1002/path.2527

[2] Palma, P., Rossi, P., Collina, G., Buccoliero, A., Ghiringhello, B., & Gilioli, E. et al. (2009). The Reproducibility of CIN Diagnoses Among Different Pathologists: Data From Histology Reviews From a Multicenter Randomized Study. American Journal Of Clinical Pathology, 132(1), 125-132. http://dx.doi.org/10.1309/ajcpbrk7d1yiuwfp

[3] The journey of a tissue biopsy - cancer testing. Ventana.com. Retrieved 5 October 2016, from http://www.ventana.com/journey/

[4] Lei He, L. Rodney Long, Sameer Antani, George R. Thoma, Histology image analysis for carcinoma detection and grading, Computer Methods and Programs in Biomedicine, Volume 107, Issue 3, September 2012, 538-556. http://dx.doi.org/10.1016/j.cmpb.2011.12.007

[5] Ting Chen, Chukka Srinivas, Group sparsity model for stain unmixing in brightfield multiplex immunohistochemistry images, Computerized Medical Imaging and Graphics, Volume 46, Part 1, December 2015, 30-39. http://dx.doi.org/10.1016/j.compmedimag.2015.04.001

[6] Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. Analytical and quantitative cytology and histology 2001; 23; 291299.

[7] Hawkins, D. (2004). The Problem of Overfitting. Journal Of Chemical Information And Computer Sciences, 44(1), 1-12. http://dx.doi.org/10.1021/ci0342472

[8] Brown, L. (1992). A survey of image registration techniques. ACM Computing Surveys, 24(4), 325-376. http://dx.doi.org/10.1145/146370.146374

[9] Zitov, B. & Flusser, J. (2003). Image registration methods: a survey. Image And Vision Computing, 21(11), 977-1000. http://dx.doi.org/10.1016/s0262-8856(03)00137-9

[10] Sarkar, A., Yuan, Q., & Srinivas, C. (2014). A robust method for inter-marker whole slide registration of digital pathology images using lines based features. 2014 IEEE 11Th International Symposium On Biomedical Imaging (ISBI).
http://dx.doi.org/10.1109/isbi.2014.6867982

[11] West, J., Fitzpatrick, J., Wang, M., Dawant, B., Maurer, C., Kessler, R., & Maciunas, R. (1999). Retrospective intermodality registration techniques for images of the head: surface-based versus volume-based. IEEE Transactions On Medical Imaging, 18(2), 144-150.
http://dx.doi.org/10.1109/42.759119

[12] Urschler, M., Kluckner, S. & Bischof, H. (2007). A framework for comparison and evaluation of nonlinear intra-subject image registration algorithms. The Insight Journal.
http://hdl.handle.net/1926/561

[13] Yin, L., Tang, L., Hamarneh, G., Celler, A., Shcherbinin, S., Fua, T. & Duzenli, C. (2009). Evaluating geometrical accuracy of image registration methods in SPECT guided radiation therapy. In Medical Image Computing and Computer-Assisted Intervention Workshop on Geometric accuracy in image guided intervention (MICCAI GAIGI), 63-70.

[14] KumarM.N, A. & S. Sheshadri, H. (2012). On the Classification of Imbalanced Datasets. International Journal Of Computer Applications, 44(8), 1-7.
http://dx.doi.org/10.5120/6280-8449

[15] Haibo He, & Garcia, E. (2009). Learning from Imbalanced Data. IEEE Transactions On Knowledge And Data Engineering, 21(9), 1263-1284.
http://dx.doi.org/10.1109/tkde.2008.239

[16] Ganganwar, V. (2012). An overview of classification algorithms for imbalanced datasets. International Journal Of Emerging Technology And Advanced Engineering, Volume 2, Issue 4, April 2012, 42-47.

[17] N. V. Chawla, K. W. Bowyer, L. O. Hall & W. P. Kegelmeyer (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, Volume 16, 321-357.

[18] Estabrooks, A., Jo, T., & Japkowicz, N. (2004). A Multiple Resampling Method for Learning from Imbalanced Data Sets. Computational Intelligence, 20(1), 18-36.
http://dx.doi.org/10.1111/j.0824-7935.2004.t01-1-00228.x

[19] Jo, T. & Japkowicz, N. (2004). Class imbalances versus small disjuncts. ACM SIGKDD Explorations Newsletter, 6(1), 40-49.
http://dx.doi.org/10.1145/1007730.1007737

[20] Guo, H. & Viktor, H. (2004). Learning from imbalanced data sets with boosting and data generation. ACM SIGKDD Explorations Newsletter, 6(1), 30-39.
http://dx.doi.org/10.1145/1007730.1007736

[21] Kubat, M. & Matwin, S. (1997). Addressing the Curse of Imbalanced Training Sets: One-Sided Selection. In Proceedings Of The Fourteenth International Conference On Machine Learning, 179-186.

[22] Weiss, G., McCarthy, K. & Zabar, B. (2007). Cost-sensitive learning vs. sampling: Which is best for handling unbalanced classes with unequal error costs?. Proc. Int. Conf. Data Mining, 35-41.

[23] Pardalos, P. (2009). Approximate dynamic programming: solving the curses of dimensionality. Optimization Methods And Software, 24(1), 155.
http://dx.doi.org/10.1080/10556780802583108

[24] Pudil, P., Novoviov, J., & Kittler, J. (1994). Floating search methods in feature selection. Pattern Recognition Letters, 15(11), 1119-1125.
http://dx.doi.org/10.1016/0167-8655(94)90127-9

[25] Jain, A. & Zongker, D. (1997). Feature selection: evaluation, application, and small sample performance. IEEE Transactions On Pattern Analysis And Machine Intelligence, 19(2), 153-158.
http://dx.doi.org/10.1109/34.574797

[26] Liu, H. & Motoda, H. (1998). Feature selection for knowledge discovery and data mining (1st ed.). Boston: Kluwer Academic Publishers.

[27] Jolliffe, I. (2007). Principal component analysis (1st ed.). New York: Springer.

[28] Hyvrinen, A. & Oja, E. (2000). Independent component analysis: algorithms and applications. Neural Networks, 13(4-5), 411-430.
http://dx.doi.org/10.1016/s0893-6080(00)00026-5

[29] Machine Learning with MATLAB (2016). Applying Supervised Learning, MATLAB ebook.
https://de.mathworks.com/

[30] Sain, S. & Vapnik, V. (1996). The Nature of Statistical Learning Theory. Technometrics, 38(4), 409.
http://dx.doi.org/10.2307/1271324

[31] Hosmer, D. & Lemeshow, S. (2000). Applied logistic regression (1st ed.). New York: Wiley.

[32] Akbani, R., Kwek, S., & Japkowicz, N. (2004). Machine Learning: ECML 2004: 15th European Conference on Machine Learning, Pisa, Italy, September 20-24, 2004. Proceedings (1st ed., pp. 39-50). Berlin, Heidelberg: Springer Berlin Heidelberg.

[33] Breiman, L. (2001). Machine Learning, 45(1), 5-32.
http://dx.doi.org/10.1023/a:1010933404324

[34] Chen, C., Liaw, A., & Breiman, L. (2004). Using Random Forest to Learn Imbalanced Data.

[35] Mittag F, Rmer M, Zell A (2015). Influence of Feature Encoding and Choice of Classifier on Disease Risk Prediction in Genome-Wide Association Studies. PLoS ONE 10(8): e0135832.
doi:10.1371/journal.pone.0135832